

GenCore version 5.1.4_p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 00:49:37 ; Search time 2859.69 Seconds
(without alignments)
13738.812 Million cell updates/sec

Title: US-09-768-781-1

Perfect score: 1350
Sequence: 1 atggacagagtttatgaat.....caaggcaaaagtgtgtctctga 1350

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sta: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sta: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_ey: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	747.2	55.3	37650	9	HSU131B10
C 2	620.8	46.0	168918	2	AC116256 Rattus no
C 3	615.4	45.6	171014	10	AL671915 Mouse DNA
C 4	441.2	32.7	1609	9	AK098608 Homo sapi
C 5	301.8	22.4	2311	10	AF155511
C 6	289.6	21.5	5096	6	AX332514 Sequence
C 7	289.6	21.5	5096	9	HSXKMTF
C 8	282.8	20.9	5215	6	AX405691 Sequence
C 9	276.4	20.5	113688	9	AC005301 Homo sapi
C 10	276.4	20.5	124823	9	AC007064 Homo sapi
C 11	249.8	18.5	168918	2	AC116256 Rattus no
C 12	218	16.1	160200	10	AC091605
C 13	218	16.1	200698	2	AL672060
C 14	215.6	16.0	166436	9	AC016752
C 15	214	15.9	161879	9	AC017005 Homo sapi
C 16	214	15.9	182083	9	AC007965 Homo sapi
C 17	209.2	15.5	37566	9	AC068541 Homo sapi
C 18	209.2	15.5	101912	9	AC000100
C 19	209.2	15.5	120951	9	AC073649 Homo sapi
C 20	209.2	15.5	145383	9	AC010682
C 21	209.2	15.5	149794	2	AC007273 Homo sapi
C 22	209.2	15.5	174082	9	AC007379 Homo sapi
C 23	194.4	14.4	175331	9	HSXKSRPXM
C 24	178.2	13.2	163495	2	AC024183
C 25	178.2	13.2	164891	2	AC022848
C 26	178.2	13.2	199174	9	AC007742 Homo sapi
C 27	171.4	12.7	1588	9	HSAP000997
C 28	164	12.1	91640	2	AC123295 Rattus no
C 29	160.2	11.9	826	6	AX079777
C 30	103.2	7.6	321	11	G03706
C 31	87.6	6.5	164891	2	AC022848
C 32	60.4	4.5	64450	9	AC073893 Homo sapi
C 33	57.8	4.3	138674	2	AC112084
C 34	51.6	3.8	125020	9	AF429315
C 35	49.8	3.7	188048	10	AL669974
C 36	46.8	3.5	68873	2	AC101531
C 37	46.8	3.5	234491	2	AL671118 Mus muscu
C 38	45.8	3.4	212374	2	AC022773 Mus muscu
C 39	45.6	3.4	144794	9	AC107055 Homo sapi
C 40	45.4	3.4	7218	6	I66494
C 41	44	3.3	152434	2	AC016262
C 42	43.4	3.2	1694	9	AF226053 Homo sapi
C 43	43.2	3.2	12229	3	AE001377
C 44	42.8	3.2	13427	6	AX346828
C 45	42.2	3.1	104455	2	AC098390 Rattus no

ALIGNMENTS

RESULT 1
HSU131B10
LOCUS
DEFINITION
Human DNA sequence from cosmid U131B10, between markers DX3366 and
DXS87 on chromosome X contains XK membrane transport protein, ESTs
and STS.
ACCESSION
273417
VERSION
X; XK membrane transport protein.
KEYWORDS
Homo sapiens
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 37650)
Dodswoth, S.
PRI 23-NOV-1999

```

TITLE      Direct Submission
JOURNAL    Submitted (17-MAY-1995) Sanger Centre, Hinxton, Cambridgeshire,
           CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk
COMMENT    requests: clonerequest@sanger.ac.uk
           IMPORTANT: This sequence is the entire insert of clone U131B10.
           The true left end of clone U131B10 is at 1 in this sequence. The
           true right end of clone U131B10 is at 37650.
           U131B10 is from the human chromosome X-specific cosmid library.
FEATURES   Location/Qualifiers
           source          1..37650
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="X"
                        /map="X"
                        /clone="LLOXNC01-131B10"
                        /clone_lib="LLOXNC01"
repeat_region 364..519
           /notes="MIR element fragment"
repeat_region 2000..2290
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
repeat_region 3347..3637
           /notes="Alu repeat: matches 1..308 of consensus"
repeat_region 3699..3800
           /notes="MSTA element fragment"
repeat_region 3712..3842
           /notes="THE1B element fragment"
repeat_region 3875..4004
           /notes="MSTA element fragment"
repeat_region 3909..4009
           /notes="THE1B element fragment"
repeat_region 4038..4085
           /notes="24 copies of 2 mer 85 & conserved"
repeat_region 4087..4378
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
misc_feature  4957..5561
           /notes="match: multiple ESTs"
misc_feature  complement(5312..5561)
           /notes="match: STS G15323"
repeat_region 5723..5842
           /notes="L1 element fragment"
repeat_region 5980..6292
           /notes="MLT2C2 element fragment"
repeat_region 5980..6048
           /notes="MLT2D element fragment"
repeat_region 6399..6759
           /notes="L1 element fragment"
repeat_region 7148..7437
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
repeat_region 7460..7754
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
repeat_region 9496..9786
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
repeat_region 10339..10630
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
repeat_region 12989..13148
           /partial
           /notes="Alu repeat: matches 308..145 of consensus"
repeat_region 13166..13450
           /notes="Alu repeat: matches 1..308 of consensus"
repeat_region 14160..14208
           /notes="L1 element fragment"
repeat_region 16486..16792
           /notes="Alu repeat: matches 1..308 of consensus"
repeat_region 17220..17477
           /partial
           /notes="Alu repeat: matches 28..308 of consensus"
repeat_region 17909..18098
           /notes="MIR3 element fragment"
           18505..18795
           /notes="Alu repeat: matches 1..308 of consensus"
repeat_region 19300..19436
           /partial
           /notes="Alu repeat: matches 1..149 of consensus"
repeat_region 19440..19644
           /partial
           /notes="Alu repeat: matches 1..224 of consensus"
repeat_region 19645..19710
           /partial
           /notes="Alu repeat: matches 242..308 of consensus"
repeat_region 19711..19870
           /partial
           /notes="Alu repeat: matches 149..308 of consensus"
repeat_region 21881..22027
           /notes="MIR element fragment"
repeat_region 23139..23303
           /partial
           /notes="Alu repeat: matches 178..1 of consensus"
repeat_region 23343..23508
           /partial
           /notes="Alu repeat: matches 308..138 of consensus"
repeat_region 23509..23800
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
repeat_region 23840..23945
           /notes="MIR2 element fragment"
repeat_region 24408..24890
           /notes="L1 element fragment"
repeat_region 24527..24554
           /notes="14 copies of 2 mer 100 & conserved"
repeat_region 24902..24947
           /notes="23 copies of 2 mer 89 & conserved"
repeat_region 24951..25431
           /notes="L1 element fragment"
           complement(25545..26171)
           /partial
           /notes="match: 232684 XK membrane transport protein"
           26196..26539
           /notes="match: 5' EST H87640 clone 252548; Paired with EST
           H87641 matching this cosmid"
           complement(26828..27189)
           /notes="match: 3' EST H87629 clone 252524"
           complement(26833..27186)
           /notes="match: 3' EST H87641 clone 252528; Paired with EST
           H87640 matching this cosmid"
repeat_region 27219..27266
           /notes="24 copies of 2 mer 96 & conserved"
repeat_region 27430..27721
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
           28995..30018
           /notes="match: multiple ESTs"
repeat_region 30042..30333
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
           30691..30724
           /notes="17 copies of 2 mer 82 & conserved"
repeat_region 30725..31015
           /partial
           /notes="Alu repeat: matches 308..1 of consensus"
           32221..32499
           /notes="Alu repeat: matches 1..301 of consensus"
repeat_region 32500..32529
           /notes="15 copies of 2 mer 100 & conserved"
repeat_region 33138..33433
           /notes="Alu repeat: matches 1..308 of consensus"
repeat_region 33557..33894
           /notes="MIR1D element fragment"
repeat_region 33914..33989
           /notes="MLT1E element fragment"
repeat_region 33935..34011

```


corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-182N4 is from the RPC1-23 Mouse PAC Library constructed by the group of Pieter de Jong.

FEATURES	SOURCE
Location/Qualifiers	1..171014
organism="Mus musculus"	/db_xref="taxon:10091"
chromosome="X"	/clone="RP23-182N4"
clone library="pBC1-23"	

Query Match	45.8%;	Score 615.4;	DB 10;	Length 171014;
Best Local Similarity	88.0%;	Pred. No. 1e-166;		
Matches 670;	Conservative 0;	Mismatches 91;	Indels 0;	Gaps 0;
QY 590	TTCCCTCGGTAGAGTTGTGCTAAATGGTATTTTCCCTGGTATCTGTCACTATGGGGCCA	649		
Db 115391	TTCTTATTTTGACGTGTGTTAATGGCTTTTCCCTGATATCTGTCACTACGGGCTA	115332		
QY 650	CCCTTTGCAATATGTTGGCTATCCAGATCAAGTACGATGACTACAAGATTCGGCTTGGGC	709		
Db 115331	CCCTGTGCAATATGTTGGCTATCCAGATCAAGTATGATGACTACAAGATTAGACTGGGC	115272		
QY 710	CACGTAGAAGTCCTTCGATCACCACATCTGGCGGACATTTGGAGATCACTTCCCGCCTCCTGA	769		
Db 115271	CACGTGAAGTCCTTTGTATCACCGTCTGGAGGACATTTGGAGATCACTTCCCGCCTCGTGA	115212		
QY 770	TTCTGGTGCTCTTCTCAGCCACATTTGAAATTTGAAGGCTGTGCCCTTCTCTAGTGCTCAACT	829		
Db 115211	TTCTGGTGCTCTTCTCGGCCACCTTGAAGTTGAAGGCTGTGCCCTTCTTAGTGCTCAACT	115152		
QY 830	TCCTGATCATCTCTTTTGAGCCCTGATTAAGTTCTGGAGAAAGTGGTGCCAGATGCCCA	889		
Db 115151	TCCTGATCATCTCTTTTGAGCCCTGGGTCAAGTTCTGGAGAGTGGTGCCCAATGCCAA	115092		
QY 890	ATAACATTGAGAAAACCTTCAGCCGGTCCGGCACTCTGGTGGTCTTGATTTCAAGTCACCA	949		
Db 115091	ACAATATTGAGAAAACCTTCAGCCGGAGTTGGCACTCTGGTGGTACTGATTTCCGTACCA	115032		
QY 950	TCCTCTATGCTGGATCAACTTCTTTGCTGGTCAGCTTTGCAAGTTGAGGTTGGCAGACA	1009		
Db 115031	TTCTCTATGCTGGATCAACTTCTCTTGTGGTCAGCCATGCAAGTTGAAAGTTGGCAGACA	114972		
QY 1010	GAGACTCTGTCGACAAAGGCGACACTGGGGACATATGGGCCCTGCACATAGTGTGAGGT	1069		
Db 114971	GAGACCTTGTTGACAAAGGTGAGAACTGGGGACATATGGGACTGCACTATAGTGTGAGGC	114912		
QY 1070	TGGTAGAAGATGTGATCATGGTCTTGTTTTTAAGTTCTTTGGAGTGAAAGTGTACTGA	1129		
Db 114911	TGGTAGAAGATGTGATCATGGTCTTGTTTTTAAGTACTTTCGGAGTCAAAGTGTACTGA	114852		
QY 1130	ATTACTGTCAATTCCTTGATTGCGCTTCGAGCTCAATTATGCTTAATCGATTTCCATTGACT	1189		
Db 114851	ATTATTGTGCAATTCCTTGATTGCGAGTCAGCTTAATTATGCTTAATCGATTTCCATTGCGG	114792		
QY 1190	TCATGCTCCCTTTCTTCCAGTACTTGATCCATTTGGCTCAGCTCTCTTCCACCCATAATGTAG	1249		

Qy	650	CCCTTGC	CAATATGTTGGCTATCCAGATCAAGTACGATGACTACAAAGATTGCCTTGGCC	709
Db	140495	CCCTTGC	CAATATGTTGGCTATCCAGATCAAGTATGATGACTACAAAGATTGCATTTGGCTC	140436
Qy	710	CAC	TAGAAGTCCTCCTCATCACATCTGGCGGACATTGGAGATCACTTCCGCGCTCCCTGA	769
Db	140435	CAGTGA	AGTCCCTTTGTATCACCATCTGGAGACATTGGAGATCACTTCCGCGCTCATGA	140376
Qy	770	TTCTG	TGCTCTTCTCAGCCACATTTGAAATTTGAAGCTGTGCCCTTCTTAGTGCTCAACT	829
Db	140375	TTCTG	TGCTCTTCTCGGCCACTTGAAGTTGAAAGCTGTGCCCTTCTTAGTGCTCACT	140316
Qy	830	TCCTG	ATCATCTCTTTTGAGCCCTGGAATTAAGTTCTGGAGAAGTGGTCCAGATGCCCA	889
Db	140315	TCCTG	ATCATCTCTTTGAGCCCTGGTTAAAGTTCTGGAGGAGTGGTCCCAATGCCAA	140256
Qy	890	ATA	CATTGGAAGAAATTCAGCCGGGTGGGCACCTCTGGTGTCTGTATTTTCAGTCAACA	949
Db	140255	ACA	ATATTGAGAAAAATTTTCAGCCGAGTTGGCAGCGTGGTGGTACTGATTTCTGTTACCA	140196
Qy	950	TCCT	TATGCTGGCATCAACTCTCTCTTGTGTGTGCTGAGTTTGCAGTTGAGTTGGCAGACA	1009
Db	140195	TCCT	TATGCTGGCATCAACTCTCTCTGCTGTGTGTGCTGAGTTTGCAGTTGAGTTGGCAGACA	140136
Qy	1010	GAG	ATCTCGTCGACAAAGGCGAGAACTGGGGACATATGGCCCTGCACATATAGTGTGAGGT	1069
Db	140135	GAG	ATCTGTGTCAGAAAGGTCAGAACTGGGGACATATGGGACATATGGGACTGACATGATGAGGT	140076
Qy	1070	TGT	TAGAGAAATGATCATGGTCTTGGTTTTTAAAGTTCTTTGAGTGAAGTGTTTACTGA	1129
Db	140075	TGT	TAGAGAAATGATCATGGTCTTGGTTTTTAAAGTCTTTTGAAGTCTTGAAGTGTTTACTGA	140016
Qy	1130	ATT	ACTGTCATTCCTTTGATTGCTTCGACGCTCATTTATGCTTATCTGATTTCCATTGACT	1189
Db	140015	ATT	ACTCCCATTCCTTTGATGGCTATTTCAGCTTATTATTTGCTTTACCTGATTTTCCATTGGAT	139956
Qy	1190	TCAT	GTCTCTTTTCTTCCAGTACTTCGATCTCCATTTGGGCTCACTCTTTCAACCCATATGATG	1249
Db	139955	TCAT	GTCTCTTTTCTTCCAGTACTTCGATCTCCATTTGGGCTCACTCTTTCAACCCATATGATG	139896
Qy	1250	TAGA	CTACCTCCATTGTGTCTGTCTCACCAGCACCCCTCGGACACGGGTTGAGAACTCAG	1309
Db	139895	TAGA	CTACCTCCATTGTGTCTGTCTCACCAGCACCCATCGGAAAGGGTTGAGAACTCAG	139836
Qy	1310	AGC	ACCTTTTGAGATGAAGCAAGCAAGTGTGTCTGA	1350
Db	139835	AG	ACATCCTGTGAGCCTTGACACAAGGCAAGTATTTGCTGA	139795

[illegible]

	Matches	747;	Conservative	0;	Mismatches	468;	Indels	15;	Gaps	1;
Qy	1	ATGCGACAGGTTTATGAAATTCCTGAGGAGCCAAATGTGGATCCGGTTTCATCTCTCGAG	60							
Db	93	ATGGAGACAGTGTTTGAGAGATGATGAAGAAACACAGGAGGAGTTTCATCTTCGAA	152							
Qy	61	GAAGATGTCTACCGTGGAGCCAAACCCCGATTTACTTTTCCATTTAGCATCTCTTTCTCC	120							
Db	153	GAAGAAATAGTCTTGGCCAGAGACTCCACTAAGCTTCTCTTTTAGCATTTATCTTCTCA	212							
Qy	121	ACCTTTTGTACTGTGGGAGGCTGCATCTGCTTTGTACATGTTTAGAATCTATTCGAAAG	180							
Db	213	ACTGTTCTCTACTGTGTGAGGTTGCTTTGGTTTATACATGTTTGAATTTATTCGAAA	272							

Db	213	ACGTGTTCTCTACGTGTGTGAGGTGCCTTTGGTTTATACATGTTGAAATTTATCGAAA	272
Qy	181	AATAGTGAAACTTACCGATGACATACACCTTTTCTTTCTTTATGTTTTCATCCATTATG	240
Db	273	GCTAAATGACACATCTCTGGATGTCATTTACCATCAGCTTTATTATTGTGGGGCCAATTTG	332
Qy	241	GTCAGTTGACCTTCATTTTGTCTCACAGAGATCTAGCCAAAGATAAACCGCTATCATT	300
Db	333	GATCAAAATTATCTGTATGTTTTTTCACAAAGACTTGTAGGAGAAATATAGGTGCTATTCTT	392
Qy	301	TTTATGCACTTAATCTCTTTGGGACCTGTTTATCAGATGTTTGTGAGGCCATGATTAAGTAC	360
Db	393	TTTGTGGCACATCTCTCTTTTAGGACCTATTGTGAGGTGTTTGACACCAATTAGAGAAATTAC	452
Qy	361	CTCACACTGTGGAGAAGAGGACGAGGAGGCCCTTATGTGACGCTTCACCCGAAAGAG	420
Db	453	CACAAATGGTGTGAAAAATCTTAAACGAGGAAGGAAGACTCAAGTTTAGTCATCACAAAG	512
Qy	421	ATGCTAATAGATGCGCAGGAGGTGCTGATAGAATTGGGAGGTGGGCCACTCCATCCGGACC	480

[illegible]

Qy	561	ATGTGGCTATCAGATCAAGTACGATGACTACAGATTCCGCTTGGGCCACTAGAAGTC	720
Db	738	ATCTGGCATCCAGATCAGCAATGATGATCTACTACCTTAAGCTACCGCTGATAGATTCT	797
Qy	721	CTCTGCATACCACTCTCTGGCGAATCTGGAGATCACTTCCCGCTCTGATTCTGTGGTCTC	780
Db	798	TTCTGTGCTGATGTGGCGTTTTTGGAGGTTATCTCAGCTGTAGTACTCTGGCAATTG	857
Qy	781	TTCTCAGGCATTTGGAATTGAAGGCTGTGGCCCTTCTTAGTGTCAACTTCTGTGATCATC	840

Qy	781	TTCTCAGCCACTTTGAATTAAGAGGCTGTGCCCTTCTPAGTGCTCAACTTCTCGTATCATC	840
Db	858	TTCATGTGATCTCTGAAGAACTGAAGAGCCTACCCGTTTGTGTTAAATCATATATTTTGTATCA	917
Qy	841	CTCTTTGAGCCCTGGATTAACTTCTGCGAAGATGGTCCAGATGCCCAATAACAATTGAG	900
Db	918	TTGTTGGCACCGTGGCTGGAGTTTGGNAAAGTGGAGCTCATCTCTCGGCAACAAGAA	977
Qy	901	AAAAAATTCAGCCGGGTGGCACTCTGGTGTCTCTGATTTTCAGTCACCATCTCTATGCT	960
Db	978	AATAAATCCAAATATGGTGGGTACAGTACTGATGCTTTTCTTGATCACACTGCTATATGCT	1037
Qy	961	GGCATCAACTTCTCTTGCTGCTCAGCTTTGCAAGTTGAGTTGGCAGCAGAGATCTCCGTC	1020
Db	1038	GCCATCAACTTCTCTGCTGCTCAGCTGAAACTGCAAGCTGTCTGATGACAAAAAATTT	1097

Db	1038	GCATCAACTCTCTCTGCTGGTCAGCACTGAAACTGCACTGTCAGATGCACAAATAATT	1097
Qy	1021	GACAAAGGGCAGAACTGGGGACATATGGGCGCTGCACCTACTATGTCAGGTTGGTAGAGAA	1080
Db	1098	GACGGGACACAGAGGTGGGGCCATAGAACTCTACACTACAGCTTTTCAGTTTTACAGAAAT	1157

QY 1081 GTGATCATGCTGGTGTGTTTAAAGTCTCTTGGAGTGAAGTGTACTGAATTAATCTGTCAT 1140
 |||||
 Db 1158 GTGATATGATATGTTATTTAGTCTCTTGGAGGGAAGAACTTGTCTGAATGTTGTGAC 1217
 |||||
 QY 1141 TCCTGTGATGCTGTCAGCTCATTAATGCTTATCTGATTTCCATTTGATGACTTCCTCTT 1200
 |||||
 Db 1218 TCATTAATGCTGTCAGCTCATATAAGCTACCTATTGGCACTGGCTTTATGCTCTC 1277
 |||||
 QY 1201 TTCTTCAGTACTTGCATCCATTCGGCTCA 1230
 |||||
 Db 1278 TTCTATCATGATTTGTACCCATGGCAGTCA 1307
 |||||

RESULT 5
 AF155511 2911 bp mRNA linear ROD 02-DEC-1999
 LOCUS Mus musculus KX antigen (Xk) mRNA, complete cds.
 DEFINITION Mus musculus KX antigen (Xk) mRNA, complete cds.
 ACCESSION AF155511 AF064772
 VERSION AF155511.1 GI:6502962
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus

REFERENCE
 1 (bases 1 to 2911)
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Collec.E., Collin.Y., Carbonnet.F., Hattab.C., Bertrand.O.,
 Cartton.J.P. and Kim.C.L.

TITLE
 Structure and expression of the mouse homologue of the Xk gene
 JOURNAL Immunogenetics 50 (1-2), 16-21 (1999)
 MEDLINE 20009522
 PUBMED 10541802

REFERENCE
 2 (bases 1 to 2911)
 Le Van Kim.C., Collec.E. and Collin.Y.
 Direct Submission
 Submitted. (01-JUN-1999) INSERM U76, INTS, Alexandre Cabanel, Paris
 75015, France

FEATURES
 Location/Qualifiers
 1..2911
 /organism="Mus musculus"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 /chromosome="X"
 /map="between Gata1 and Cybb"
 /sex="male"
 /tissue_types="skeletal muscle"
 /dev_stage="8-12 weeks"
 1..2911
 /gene="Xk"
 1..141
 /gene="Xk"
 142..1482
 /gene="Xk"
 /notes="Homo sapiens Kx blood group protein homolog"
 /codon_start=1
 /product="KX antigen"
 /protein_id="AAPI4527.1"
 /db_xref="GI:6502963"
 /translation="MKPPASVIASVFLVFAETAALYLSSTYRSAGDRMQVLTLLFS
 LMPCLVQFTLLFVHRLSDRPLALMLHLQLGPLYRCCEVFCIYQSDNEEYVS
 ITKRQPKDGLSEVEKEVGOAEGKLIHRSAPSRASVIOAFIGSAPOLTLQIYTV
 LEONITGRCFIMTSLLSIVYGNLRNLIKIKYDEYEVKUYKPLAYVUCIFLWRSFE
 IATRVIVLFTSLVKIWKVAVLVNFFSLFPIWVFCWGSFPFENIEKLSRVGT
 TIVLCFTULIYAGINMPCVSAQVLKIDNPILSKQNWYRLLIYYTRFIENSVLILL
 WYFKTDIYVYCAPLLILQLLIGCTGILFVYQFPHPKKLFSSSVSESFRALL
 RCACWSLSLRKSEPVGRIDTLKACTEQDVMPTTSKVIPEATDITWAVDLCSA"
 1483..>2911
 /gene="Xk"

BASE COUNT 741 a 700 c 609 g 861 t
 ORIGIN

Query Match 22.4%; Score 301.8; DB 10; Length 2911;
 Best Local Similarity 55.5%; Pred. No. 5.7e-76;

Matches 629; Conservative 0; Mismatches 492; Indels 12; Gaps 2;
 QY 118 TCACACCTTTTGTACTGTGGGAGGCTGCATCTGCTTGTACATGTTAGAAATCTATCGA 177
 |||||
 Db 169 TCCGTGTTCTGTTGTTGTCGCGGAGACGCGCGGCTCTACTTGTGAGCAGACCTACGCG 228
 |||||
 QY 178 AAGAATAGTGAACCTTACCAGATGACATACACCTTTCTTCTTCTTATGTTTTCATCCATT 237
 |||||
 Db 229 TCCGACGGGACCCGATGTGACAGTGCTGACGCTCTTCTCTCTGATGCCCTCGGCC 288
 |||||
 QY 238 ATGCTCCAGTTGACCTCATTTTGTCCACAGAGATCTAGCCAAAGATAAACCCTATCA 297
 |||||
 Db 289 CTGGTGAGTTTCAGCTCTCTCTGTCACCGACCTCAGCGCGATCGCCCACTGGCG 348
 |||||
 QY 298 TTATTTATGATCTAATCTCTTGGGACCTGTTATCAGATGTTTGGAGGCCATGATTAAG 357
 |||||
 Db 349 CTGCTCATGACCTGCTCCAGCTCGGCCCTGTACAGGTGTGTGAAGTCTTTTGTATC 408
 |||||
 QY 358 TACCTCACACTGTGGAAGAAAGAGGAGGAGGAGGAGCCCTATGTGAGCTCACCCGAAA- 416
 |||||
 Db 409 TAC-----TGTCACTCCGACCCAGATGAAGAACCTTATGTGAGCATCACTAAGAAA 459
 |||||
 QY 417 --GAAGATGCTAATAGATGGGAGGAGTGCTGATAGAAATGGAGGTGGGCCACTCCATC 474
 |||||
 Db 460 CGACAGATGCCAAAGATGGCTTTTCAGAGGAGGTGGAGAAAGAGTGTGGCCAGCAGAA 519
 |||||
 QY 475 CGGACCTTGGTATGACCCGCAATGCCCTAAGAGTATGTACAGATCAAGCCTTCTCTG 534
 |||||
 Db 520 GGCAAGCTGATTAATCCACCGTTCTGCAATTCAGCCGGGATCAGTGATACAGGCTTTCCTA 579
 |||||
 QY 535 GGCTCAGTGCCCGGAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGCAGAGGTTCCC 594
 |||||
 Db 580 GGCTTGGCCCAACCTGACCTAACCTGATATACTGCTTGGAGCAGAACATCACT 639
 |||||
 QY 595 CTGGGTAGAGTTGTGCTAATGTTTCTCCCTGCTATCTGCACCTATGGGCCACCCCTT 654
 |||||
 Db 640 ACTGGAAGATGCTTTCATCATGACCTTGTCCCTGCTCTATTGTCTATGGCGCTTACGT 699
 |||||
 QY 655 TGAATATGTTGGTATTCAGATCAAGTACGATGACTCAAGATTCGGCTTGGGCCACTA 714
 |||||
 Db 700 TGCAACATCTTAGCCATCAAAATCAAGTATGATGATGAGGTCAAAAGTCAAAACCCCTG 759
 |||||
 QY 715 GAAGTCTCTGCTCATTACCATCTGGCGGACATTGGAGATCACTTCCCGCTCTCTGATCTG 774
 |||||
 Db 760 GCTTATGCTGTATCTCTCTCGGAGAGCTTGGAGATGGCACCCGAGTCACTGCTCTG 819
 |||||
 QY 775 GTGCTCTTCTCAGCCACTTTGAAATTTGAAGGCTGTGCCCTTCTAGTGTCTCAACTTCTCTG 834
 |||||
 Db 820 GTCTCTTTTACCTCTGCTCTGAAGATCTGGGTGGTGGCAGTCATCTGCTCAACTTCTTC 879
 |||||
 QY 835 ATCATCTCTTTGAGCCCTGGATTAAGTTCTGGAGAAAGTGTGGCCAGATGCCCAATTAAC 894
 |||||
 Db 880 AGCTTCTTCTTATATCTCTGATCGTTTCTGTTGTCAGTGGCTCCCATTCCTCTGAGAAC 939
 |||||
 QY 895 ATTGAGAAAACCTTCAGCCCGGTCGGCACTCTGTTGGTGGCTGATTTCACTACCACTCTC 954
 |||||
 Db 940 ATAGAGAGGCCCTTAAGTAGGGTGGGTACCACTCGTGTGCTCTGCTCTCTCACTTTACTC 999
 |||||
 QY 955 TATCTCGCATCAACTCTCTTGTGCTGTGCTGAGTTTGCAGTTTGGAGTTTGGCAGACAGAT 1014
 |||||
 Db 1000 TATGCGGCATCAACATGTTCTGCTGTGCTGAGTGTTCAGCTGAAATCGACATCCGGAG 1059
 |||||
 QY 1015 CTGCTGACAAAGGGCAGAACTGGGGACATATGGCCCTGCACATPATGTGTGAGGTGGTA 1074
 |||||
 Db 1060 CTATCAGCAAGTCCCAAGAACTGGTACCGCTGCTCATTTACTATACAGAGATTCATC 1119
 |||||
 QY 1075 GAGATGTGATCATGCTGTTGTTTAAAGTCTTCTTGGAGTGAAGTGTACTGAATTAC 1134
 |||||
 Db 1120 GAGAACTCCGCTCTCTGCTCTGCTGTTATTTTAAACATGACATATACATGTATGTG 1179
 |||||
 QY 1135 TGTCATCTCTGATGCTTCAGCTCATTTATGCTTATCTGATTTCCATTTGACTTCAATG 1194
 |||||
 Db 1180 TGTGCGGCTCTGTTGATCTTGCAGCTGCTTATTTGGGTACTGCATCGGCACTTCTTCTATG 1239
 |||||


```

y 1195 CTCCTTTCTTCCAGTACTGTCATCTGCGCTCAGCTTCCACCATTAATGT 1247
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 1240 CTTGTGTTCTATCAGTCTTTTCCACCTTGCAAAAAGCTCTTCTCTCCAGTGT 1292
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
X332514
OCUS AX332514 5096 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3023 from Patent WO0194629.
ACCESSION AX332514
VERSION AX332514.1 GI:18123148
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrikan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
Gene sets
JOURNAL Patent: WO 0194629-A 3023 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Location/Qualifiers
1..5096
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1392 a 1064 c 1022 g 1618 t.
ORIGIN
Query Match 21.5%; Score 289.6; DB 6; Length 5096;
Best Local Similarity 54.6%; Pred. No. 2.2e-72;
Matches 627; Conservative 0; Mismatches 509; Indels 12; Gaps 2;

y 118 TCACCTTTTCTACTGTGGGAGGCTGCATCTCTGTTGTACATGTTAGATCTATCGA 177
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 110 TCCGTGTTCTGTTGCGGCCGAGACAACGGGGCGCTCAGCTGAGCAGCACCTACCGC 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 178 AAGATATAGTGAACATTACCGGATGACATACACCTTTTCTTTCTTTATGTTTTCATCCATT 237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 170 TCGGGGGGGGACCGCATGTGGCAGGCGCTGACGTGCTTTCTCTGCTACTGCTCGGG 229
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 238 ATGGTCCAGTTCACCTCATTTTGTCCACAGAGTCTAGCCAAAGATAAACCGCTATCA 297
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 230 CTCGTGACCTCAGCTTCTCTCTGACACCGGACCTCAGCGCGACCGCGCTCGTA 289
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 298 TTATTTATGATCTAATCTCTTGGGACCTGTTTATCAGATGTTTGGAGGCCATGATTAAG 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 290 CTGCTGCTGCACCTGCTGCAACTTGGGCCCCCTTTTTCAGGTGTTTGAAGTCTTCTGCATC 349
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 358 TACCTCACACTGTGGAGAAAGAGGAGCGAGGAGCCCTATGTCAGCTCACCAGAAAG 417
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 350 TACTTTC-----AGTCAGGCAACAATGAAGAGCCTTATGTCAGTATCACCAAGAAG 400
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 418 AAG---ATGCTAATAGTCGGAGGAGTGTCTATAGATGGGAGGTGGGCCACTCCATC 474
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 401 AGGCAAAATGCCAAAATAATGCTCTCAGAGGAGATTGAGAAGAGGTGGGCCAGGAGAA 450
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 475 CGGACCTGGCTATGCACCGCAATGCTTACAAACGTATGTACAGATCCAAAGCTTCCCTG 534
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 461 GGCAACTAATCACCCACCGATCAGCGTTTCAGCGGGCGTGGTGATCCAGGCTTCTTG 520
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 535 GGTCTAGTCCCGCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGAGAGGTTCC 594
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 521 GGTCTAGCCTCCCGCAGCTGACCTTACAGCTGTACATAAGTGTATGTCAGCAGGAGCGTCACT 580
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 595 CTGGGTAGAGTGTGTATGTTATTTCCCTGGTATCTGTACCTATGGGCGCACCTT 654
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 581 GTTGGGAAGAGTCTCTCTATGACCATCTCTGTTGCTTATGTTGAGGCTTGGCG 640
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 655 TGAATATGTTGGCTATCCAGATCAAGTACGATGACTACAAAGATTGCGCTTGGGCGCACTA 714
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 641 TGAACATCTAGCCATCAAAATCAAGTACGATGAGTATGAAGTCAAAAGTGAAGCCTCTG 700
Qy 715 GAAGTCTCTGCATCACCATCTGGGGACATTTGGAGATCACTTCCCGCTCTCTGATTTTG 774
Db 701 GCCTATGTCTGATCTTCTCTGTGGAGGAGCTTTGAGATTTGCCACTCGAGTTGTAGTCTG 760
Qy 775 GTGCTCTTCTCAGCCACTTTGAAATTTGAAGCTGTGCGCTCTCTCTAGTGTCAACTTCTCTG 834
Db 761 GTCTCTTTTACTCTCGTCTCTGAGACCTGGGTGGTGTATATACTCATCACTTCTTCTC 820
Qy 835 ATCAPCTCTTTTGAGCCCTGGAATTAAGTTCTGGAGAAATGTGGCCAGATGCCCAATTAAC 894
Db 821 AGTTTCTTCTTGTACCTCCCTGGATCTCTTCTGTGGTGCAGTGGTTCCTCCCATTCCTCGAGAAC 880
Qy 895 ATTGAGAAAACCTTCAGCGGGTGGGCTCGGCACTCTGGTGTCTCTGATTTTCACTCAGTCACCATCTC 954
Db 881 ATAGAGAAGGCCCTCAGTAGAGTGGGCAACCACTTGTACTATGCTTTTCTAACTTTTACTC 940
Qy 955 TATGTGGCATCAACTTCTCTTCTGTGTGTCAGCTTTTGCAGTTGAGGTGGCAGACAGAGAT 1014
Db 941 TATCTGTGTATCAACATGTTCTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1000
Qy 1015 CTCGTGCAAAAGGCGCAACTGTGGGACATATGGGCTGTGCACTATAGTGTGAGGTGTGTA 1074
Db 1001 CTCATCAGCAAGTCCCAATAATTGTGCCAGCTACTGTGTGTATTTACATGATAAGATTCTATC 1060
Qy 1075 GAGATGTGATCATGCTTCTGTTTAAAGTTCTTGTGAGTGAAGTGTACTGAATTAAC 1134
Db 1061 GAGATGTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1120
Qy 1135 TGTCTATCTCTGATTGCTTGCAGCTCATTTATGCTTATCTGATTTTCCATTTGACTTCTCATG 1194
Db 1121 TCGGCACCTCTGTGTGTCTGTGAGCTGTCTATTTGGGTACTGTCAGACCACTTCTTCTCATG 1180
Qy 1195 CTCCTTTTCTTCCAGTACTTGCATCCATTTGGCTCCTCTCTTCCACCATTAATGATAGAC 1254
Db 1181 CTGTATTTCTATCAGTCTTCTTCCACCTTTCGAAAAAGCTTTTCTTCCAGTGTCTTCTGAA 1240
Qy 1255 TACCTCCA 1262
Db 1241 GGCTTTCA 1248

```

RESULT 7

```

HSXKMT 5096 bp mRNA linear PRI 19-MAY-1999
LOCUS HSXKMT Homo sapiens mRNA for membrane transport protein (XK gene).
DEFINITION Homo sapiens mRNA for membrane transport protein (XK gene).
ACCESSION Z32684
VERSION Z32684.2 GI:4883432
KEYWORDS membrane transport protein; XK gene.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 5096)
AUTHORS Ho, M., Chelly, J., Carter, N., Danek, A., Crocker, P. and Monaco, A.P.
TITLE Isolation of the gene for McLeod syndrome that encodes a novel
membrane transport protein
JOURNAL Cell 77 (6), 869-880 (1994)
MEDLINE 94273191
PUBMED 8004674
REFERENCE
2 (bases 1 to 5096)
AUTHORS Ho, M.F.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1994) Meng F Ho, Human Genetics, Imperial Cancer
Research Fund, Institute of Molecular Medicine John Radcliffe
Hospital, Headington, Oxford, OXON, OX3 9DU, United Kingdom
REMARK revised by [3]
REFERENCE
3 (bases 1 to 5096)
AUTHORS Ho, M.F.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1997) Meng F Ho, Human Genetics, Imperial Cancer
Research Fund, Institute of Molecular Medicine John Radcliffe

```


Hospital Headington, Oxford, OXON, OX3 9DU, United Kingdom
On May 20, 1999 this sequence version replaced gi:2570027.

FEATURES
Source
Location/Qualifiers
1. 5096
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="XK"
83..5068
/genes="XK"
83..1417
/genes="XK"
/citation=[1]
/codon_start=1
/product="membrane transport protein"
/protein_id="CAA83632.2"
/db_xref="GI:4883433"
/translations="MKPPASVLASVFLVAETTTAALSLSTVRSRGGDRMWAQLTLFLS
LIPCALVOLTLLEVRDLSDRLDRLVLLHLLQLPLFRCEFPVCIYFQGNNEPVS
ITKRPNGLSSEEIEKVGQAEGLITHRSAFSAIVQAPLPSAPQTLQLYISV
MODVTGVSLSMTISLSIVGALRCNLIAIKIKYDEYEVKPKLAYVICIFLRSFE
IATRVVLFTSVLTKVWVILINFPFPLPFCWSPSPFENIEKLSRVGT
TIVLCFLTLTYCNMFCWSAVOLKIDSPDLISKSHWVQLLYVMIRFENAILLL
WYLFKTDIYMYVCAPLLVLQLLIGYCTAILFMLVYQFHPCKKLPSSSVSGFQRL
RCFCWACRQKPCPEPIGKEDLQSSRDRETPSSKTSPEPGQFLNADLCSA"
83..327
/genes="XK"
/number=1
328..590
/genes="XK"
/number=2
591..1417
/genes="XK"
/number=3
5036..5041
/genes="XK"
5063..5068
/genes="XK"
5092
polyA_site
1392 a 1064 c 1022 g 1618 t
ORIGIN
Query Match 21.5%; Score 289.6; DB 9; Length 5096;
Best Local Similarity 54.6%; Pred. No. 2.2e-72;
Matches 627; Conservative 0; Mismatches 509; Indels 12; Gaps 2;
Qy 118 TCACCTTTTGTACTGTGGGAGGCTGCATCTGCTTTGTACATGGTTAGAACTATCGA 177
Db 110 TCGGTGTTCTGTTGCGGCGAGACAACGGCGCGCTCAGCTGACGACCACTACCGC 169
Qy 178 AAGAAATAGTAACTTACCGATGACATACACCTTTTCTTTTATGTTTTCATCCATT 237
Db 170 TCGGGCGGGACCGCATGTGGCAGCGCTGACGTTGCTTTCTCGTACTGCTTGGCG 229
Qy 238 ATGTCAGTTGACCTTATTTTGTCCAGAGATCTAGCCAAAGATAACCGCTATCA 297
Db 230 CTCGTGAGCTCAGCTTCTCTTCTGATACCGGACCTCAGCCGCGCCGCTCGTA 289
Qy 298 TTAATTTATGATCAATCTCTTGGGACCTGTTATCAGATGTTTGGAGGCGCATGTAAG 357
Db 290 CTGCTGCTGACCTGCTGCACTTGGGCGCCCTTTTCAGGTGTTTGAAGCTTCTGCATC 349
Qy 358 TACCTCACTGTGGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 417
Db 350 TACTTTC-----AGTCAGGCAACAATGAAGAGGCTTATGTCAGTATCACCAGAAAG 400
Qy 418 AAG---ATGCTAATAGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 474
Db 401 AGGCAATGCCAAAAATGGCTCTCAGAGGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGAA 460
Qy 475 CGGACCTGGCTATGACCGCAATGCTTACAAAGTATGTCACAGATCCAGGCTTCTG 534
Db 461 GGCAATATATCACCACCGATCAGCGTTACGCCGCGCTGCGTATCCAGGCTTCTTGT 520
Qy 535 GGCTAGTGCCCGCAGCTGACCTATCAGCTCTATGTAGCCTGATCTCTGCAGAGGTTCC 594
Db 521 GGCTAGCGCCCGCAGCTGACCTTACAGCTGTACATAAGTGTGATGACGAGGAGGCTACT 580
Qy 595 CTGGGTAGAGTTGCTTAATGGTATTTTCCCTGTGTCTGCTACCTATGAGGAGGAGGAGG 654
Db 581 GTTGGGAAGAAGTCTCTCTCATGACCATATCCCTGTGTGCTTGTGTATGAGGAGCTTGGCG 640
Qy 655 TGCAATATGTTGGCTATCCAGATCAAGTACGATGACTACAGATTCGGCTTTGGGCGACTA 714
Db 641 TGCAATATGTTGGCTATCCAGATCAAGTACGATGACTACAGTCAAGTCAAGTCAAGCTCTG 700
Qy 715 GAAGTCTCTGCTACACCTCTGGCGACATTTGGAGATCACTTCCCGCTCTCTGATCTG 774
Db 701 GCCTATGCTGTATCTCTCTGTGGAGGAGCTTTGAGATTCGCACTCGAGTTGTAGTCTG 760
Qy 775 GTGCTCTTCTCAGCCACTTTGAAATGGAAGCTGTGGCTTCCCTTCTAGTGTCTCAACTTCTC 834
Db 761 GTCTCTTTTACCTCCGCTCTGGAAGACCTGGGTGGTGTATTAATACTCATCAACTTCTTC 820
Qy 835 ATCATCTCTTTTACCTCCGCTCTGGAAGACCTGGGTGGTGTATTAATACTCATCAACTTCTTC 894
Db 821 AGTTCTTCTTGTACCTCGATCTCTTCTGGTGAGTGGTTCCTCCCACTTCCCTGAGAAC 880
Qy 895 ATTGAGAAAACTTACGCGGCTGGCACTCTGGTGGTCTGATTTTCACTGATCAGTCCCTC 954
Db 881 ATAGAGAAAGGCGCTCAGTACGATGGGCACTTGTACTATGCTTTTCTTAACCTTACTC 940
Qy 955 TATGCTGGCATCACTTCTCTGCTGCTGATTTGAGTGGAGTGGAGTGGAGTGGAGTGGAG 1014
Db 941 TATCTGTTATCAACATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1000
Qy 1015 CTGCTGCAAAAGGCGAGAACTGGGACATATGGGCTGCACTATAGTGTGAGGTTGGTA 1074
Db 1001 CTATCAGCAAGTCCCAATATGGTACGACTAGTGTGATTAATGATGATGATGATGATG 1060
Qy 1075 GAGATGTGATCATGCTTCTGTTTAAAGTCTTCTGGAGTGAAGTGTGTTACTGATTAAC 1134
Db 1061 GAGATGCACT 1120
Qy 1135 TGTCATCTCTGATGCTTGGCTTGCAGCTCATTTATGCTTATCTGATTTCCATGATTTGATG 1194
Db 1121 TGGCAGCTCTGTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1180
Qy 1195 CTCTTTTCTTCCAGTACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1254
Db 1181 CTGTATTTCTATCAGTCTTCCACCTTGCACCTTGCACCTTGCACCTTGCACCTTGCAC 1240
Qy 1255 TACCTCA 1262
Db 1241 GGCTTCA 1248
RESULT 8
AX405691
LOCUS
DEFINITION
Sequence 106 from Patent WO0222660.
ACCESSION
AX405691
VERSION
AX405691.1 GI:21438821
KEYWORDS
human.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1
Tang, Y. T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q. A., Ren, F.,
Xue, A. J., Yang, Y., Wehrman, T. and Drmanac, R. T.
TITLE
Novel nucleic acids and polypeptides
JOURNAL
Patent: WO 0222660-A 106 21-MAR-2002;
HYSEQ, INC. (US)
FEATURES
Location/Qualifiers
Source
1. 5215

```
/organism="Homo sapiens"
/db xref="taxon:9606"
304..1536
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD34798.1"
/db xref="GI:21438822"
/translation="MMQALTLLSLLPCLALVQLTLLFVHRDLRDRPLVLLHLLQAG
PLRFCEVFCIFQSGNNEPYSITKKRQMPKNGLEIEKEVQGAQGLIHLIRSAF
KDEYEVKPLAVYCIPLARSPEIATRVVVLVLFPSVLTWVVLILINFFSEFLYP
WILFWGSGSPFENIEKALSRGTIVLCELTLLYGINPFCMSAVOLKIDSDPLISK
SHNYQLVYIMRFIENAILLLVLFKTDIYMYVCAPIVLQLLIGYCTAILFMLV
FYQFFHPCKKLFSSSVSEGGFQRLRCFCMACROQKPCPEIKGKDLQSSRDRDETPSS
KTSPEQOFLNADLCSA"
CDS
1458 a 1058 c 1035 g 1664 t
KIGIN
Query Match 20.9%; Score 282.8; DB 6; Length 5215;
Best Local Similarity 55.7%; Pred. No. 2e-70;
Matches 589; Conservative 0; Mismatches 457; Indels 12; Gaps 2;
y 208 ACCTTTCTTTCTTTATGTTTTCATCTATGTCAGTTCACCCCTCATTTTGTCCAC 267
b 319 ACCTTTCTTTCTTTATGTTTTCATCTATGTCAGTTCACCCCTCATTTTGTCCAC 378
y 268 AGAGATCTAGGCAAGATAAACCGCTATCATTTATGTCATCTAATCTCTTGGGACCT 327
b 379 CGCACCTCAGCGCGGACCGCGCTCTACTGCTGTCACCTGCTGCAACTTGGGCC 438
y 328 GTTATCAGATCTTTGGAGGCCATGATTAAGTACCTTCACACTGTGGAAGAAAGAGAGAG 387
b 439 CTTTTCAGGTGTTTGAAGTCTTCTGCATCTACTTTC-----AGTCAGGCAACAAT 489
y 388 GAGAGCCCTATGTCAGCTCACCGCAAGAAG-----ATGCTATAGATGGCGAGAGGTG 444
b 490 GAAGAGCCCTATGTCAGTATCACCAAGAAGAGGCAAAATGCCCAAAATTTGGCTCTCAGAG 549
y 445 CTGATAGATGGGAGTGGGCGACCTCCATCCGACCTTGGCTATGTCACCGCAATGCCTAC 504
b 550 GAGATTGAGAGAGGTGGCGCAGCAGAGCAAACTAATCACCACCGATCAGCGTTC 609
y 505 AAACGTATGTACAGATCAAGCCCTTCTGGGCTCAGTGGCCCGACGACCTATCAGCTC 564
b 610 AGCGGGGCTCGGTGATCCAGGCTTCTTGGGCTCAGCGCCCGCAGCTGACCTACAGCTG 669
y 565 TATGTAGCCTGATCTTCGAGAGGTTCCTTGGGTAGTGTGCTATGCTATGTTATTTCC 624
b 670 TACATAAGTGTATGACGAGGACGTCATGTTGGAAGAAGTCTCTCATGACCATATCC 729
y 625 CTGGTATCTGTACCTATGGGGCCACCCCTTTGCAATATGTTGGCTATCCAGATCAAGTAC 684
b 730 CTGTTGTCATGTTGATGAGGCTTGGCTGCAACATCTTAGCCATCAAAATCAAGTAC 789
y 685 GATGACTACAAGATTCGGCTTGGGCGACCTAGAAAGTCTCTGTCATCACCATCTGGCGGACA 744
b 790 GATGAGTATGAAGTCAAAAGTGAAGCTCTGGGCTATGCTGATCTTCTGTTGGAGGAGC 849
y 745 TTGGAGATCACTTCGGGCTCTGATTTCTGGTGTCTTCTCAGCCACTTTGAAATGAAG 804
b 850 TTTGAGATTTGGCACTCGAGTGTAGTCTGGTCTCTTTACCTCCGCTGGAAGACCTGG 909
y 805 GCTGTGCTCTTCTAGTGTCAACTTCTTCTGATCATCTCTTTGAGCCCTGGATTAAGTTC 864
b 910 GTGGTGGTATTAATCTATCACTTCTTCTAGTCTTCTTGTACCCCTGGATCCTCTTC 969
y 865 TGGAGAAGTGTGCGCCAGATGCGCAATTAACATTGAGAAAAAATTCAGCCGGTGGCACT 924
b 970 TGGTCAGTGGTTCCTCCATTCCTGAGAACATAGAGAAGCCCTCAGTAGAGTGGGACCC 1029
y 925 CTGGTGGTCTGATTTTCTAGTCAACATCTCTATGCTGGGATCAACTTCTTCTGCTGGTCA 984
b 1030 ACCATTGACTATGCTTTCTAACTTTTACTCTTACTGATCAACATGTTCTGCTGGTCT 1089
```

```
QY 985 GCTTTGAGTTGAGTTGGCAGACAGAGATCTCGTCGACAAAGGGCAGAACTGGGGACAT 1044
DB 1090 GCTGTACAGCTGAAATTTGACAGCCCTGACCTCATCAGCAAGTCCCATTAATTTGTACCAG 1149
QY 1045 ATGGGCTGCACTATAGTGTGAGTTGGTAGAAGATGTGATCATGTTCTTGGTTTAAAG 1104
DB 1150 CTACTGTGTATTACATGATAAGATTTCATCGAAGATCCATCTCTCTCTCTCTGTGTGTAT 1209
QY 1105 TTTCTTGGAGTGAAGTTTACTGAAATTTACTGTCTATTCCTTGTGTTGCTTGCAGCTCAT 1164
DB 1210 CTTTTCAGACTGACATCTATATGATGTGTGGCAGCTCTGTGTGGTCTCTGCAGCTCTC 1269
QY 1165 ATTCTTATCTGATTTCCATGACTTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1224
DB 1270 ATTGGTACTGCACAGCAATCTCTCTCATGCTGTGATTTCTATCATGTTCTTCCACCTTGC 1329
QY 1225 CGTCACCTTCCACCATATGATGATAGTACTACCTCCA 1262
DB 1330 AAAAAGCTCTTTCTTCCAGTGTCTTCTGNAAGGCTTTCA 1367
```

```
RESULT 9
AC005301/c 113688 bp DNA linear PRI 27-MAY-2000
LOCUS Homo sapiens chromosome 22q11 clone p15j16, complete sequence.
DEFINITION AC005301
ACCESSION AC005301
VERSION AC005301.22 GI:7107554
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 113688)
AUTHORS Cantu, L.A., McDermid, H. and Roe, B.A.
JOURNAL Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 113688)
REFERENCE 2 (bases 1 to 113688)
AUTHORS Cantu, L.A., McDermid, H. and Roe, B.A.
JOURNAL Direct Submission
TITLE 2 (bases 1 to 113688)
AUTHORS Cantu, L.A., McDermid, H. and Roe, B.A.
JOURNAL Submitted (18-JUL-1998) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 3 (bases 1 to 113688)
AUTHORS Cantu, L.A., McDermid, H. and Roe, B.A.
JOURNAL Direct Submission
TITLE 4 (bases 1 to 113688)
AUTHORS Cantu, L.A., McDermid, H. and Roe, B.A.
JOURNAL Submitted (28-FEB-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 5 (bases 1 to 113688)
AUTHORS Cantu, L.A., McDermid, H. and Roe, B.A.
JOURNAL Direct Submission
TITLE 6 (bases 1 to 113688)
AUTHORS Cantu, L.A., McDermid, H. and Roe, B.A.
JOURNAL Submitted (25-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 6 (bases 1 to 113688)
AUTHORS Cantu, L.A., McDermid, H. and Roe, B.A.
JOURNAL Direct Submission
TITLE COMMENT
On Feb 28, 2000 this sequence version replaced gi:6056215.
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below.
AC005301 (p15j16) 77414 113688 (0) overlaps AC007064 (p8708) 1 36275
(88548).
```

```
FEATURES          Location/Qualifiers
Source            1. 113688
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="22q11"
                  /clone="p15j16"
BASE COUNT       37158 a 22385 c 21889 g 32256 t
ORIGIN
Query Match      20.5%; Score 276.4; DB 9; Length 113688;
Best Local Similarity 64.9%; Pred. No. 2.4e-68;
Matches 409; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
QY 601 AGAGTTGTGCTAATGTTATTTCCCTGGTATCTGTCTACCTATGGGCCACCCCTTGCAT 660
Db 101529 ATAGCATGCTGATGACATTTTCCCTGTTATCAGTTACTTATGGGCCCATTCGCTGCAAT 101470
QY 661 ATGTGTGCTATCCAGATCAAGTAGCTAGCTACAAAGATTGCGCTTGGGCCACTAGAAGTC 720
Db 101469 ATACTGCCATCCAGATCAGCAATGATGATACTACCAATTAAGCTACCGCCGATAGAATTC 101410
QY 721 CTCTGCATCACCATCTGGCGGACATTGGAGATCACTTCCGCGCTCTGTGATCTGTGGCTC 780
Db 101409 TTCTGTGCTGATGTGGCGTTTTTTGGAGGTTATCTCAAGTGTAGTGAATCTGGCAATTT 101350
QY 781 TTCTCAGCCACTTTGAATTAAGGCTGTGCCCTTCCTCTAGTGTCAACTTCTCTGATCATC 840
Db 101349 TTCAATGCATCTCTGAAACTGAAGAGCGCTACCCGTTTTTGTATCATATATTTGTATCA 101290
QY 841 CTCTTTGAGCCCTGGATTAAGTTCTCGAGAAGTGGTCCAGATGCCCAATAACATTTGAG 900
Db 101289 TTGTTGACCGTGGCTGGAGTTTGGAAAGTGGAGCTCATCTTCTGGCAACAAGAA 101230
QY 901 AAAAACTTCAGCGGGTGGGCACTCTGGGTGCTGTGATTTCAAGTCAACATCTCTATGCT 960
Db 101229 AATAAATTCGAATATGGTGGGTACAGTACTGATGCTTTCTTGTGATCACTGCTATATGCT 101170
QY 961 GGATCAACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 101169 GCATCAACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101110
QY 1021 GACAAAGGGGAGAACTGGGACATATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 101109 GACGGGAGACAGAGTGGGCGCATAGATCTCTACATACAGCTTTCAGTTTTAGAAAT 101050
QY 1081 GTGATCATGTCTTGTGTTTAAAGTCTTTGGAGTGAAGTGTACTGAATTAAGTGTGCTAT 1140
Db 101049 GTGATAATGATATGTTATGTTAGTCTTTGGAGGGAAGAACTTGTCTGAATTTGTTGAC 100990
QY 1141 TCCTTGATTCCTTGGAGCTCATTTATGCTTATCTGATTTCCATTGATTCATGCTCCTT 1200
Db 100989 TCATTAATTCGGTGGAGCTCATCATAGCTATCTATTGGCCACTGGCTTTATGCTCCTC 100930
QY 1201 TTCTTCCAGTACTTGCATCCATTCGCTCA 1230
Db 100929 TTCTATCAGTATTTGTACCAATGCGACTCA 100900

FEATURES          Location/Qualifiers
Source            1. 124823
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="22q11"
                  /clone="p8708"
BASE COUNT       41074 a 23932 c 23324 g 36493 t
ORIGIN
Query Match      20.5%; Score 276.4; DB 9; Length 124823;
Best Local Similarity 64.9%; Pred. No. 2.4e-68;
Matches 409; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
QY 601 AGAGTTGTGCTAATGTTATTTCCCTGGTATCTGTCTACCTATGGGCCACCCCTTGCAT 660
Db 24116 ATAGCATGCTGATGACATTTTCCCTGTTATCAGTTACTTATGGGCCCATTCGCTGCAAT 24057

JOURNAL
REFERENCE         2 (bases 1 to 124823)
AUTHORS           Ray,L.A., Loh,P., Qi,S., Sloan,D., McDermid,H. and Roe,B.A.
TITLE             Direct Submission
JOURNAL           Submitted (11-MAR-1999) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE         3 (bases 1 to 124823)
AUTHORS           Ray,L.A., Loh,P., Qi,S., Sloan,D., McDermid,H. and Roe,B.A.
TITLE             Direct Submission
JOURNAL           Submitted (26-OCT-1999) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE         4 (bases 1 to 124823)
AUTHORS           Ray,L.A., Loh,P., Qi,S., Sloan,D., McDermid,H. and Roe,B.A.
TITLE             Direct Submission
JOURNAL           Submitted (27-OCT-1999) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE         5 (bases 1 to 124823)
AUTHORS           Ray,L.A., Loh,P., Qi,S., Sloan,D., McDermid,H. and Roe,B.A.
TITLE             Direct Submission
JOURNAL           Submitted (05-NOV-1999) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE         6 (bases 1 to 124823)
AUTHORS           Ray,L.A., Loh,P., Qi,S., Sloan,D., McDermid,H. and Roe,B.A.
TITLE             Direct Submission
JOURNAL           Submitted (06-NOV-1999) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE         7 (bases 1 to 124823)
AUTHORS           Ray,L.A., Loh,P., Qi,S., Sloan,D., McDermid,H. and Roe,B.A.
TITLE             Direct Submission
JOURNAL           Submitted (29-FEB-2000) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE         8 (bases 1 to 124823)
AUTHORS           Ray,L.A., Loh,P., Qi,S., Sloan,D., McDermid,H. and Roe,B.A.
TITLE             Direct Submission
JOURNAL           Submitted (25-MAY-2000) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE         9 (bases 1 to 124823)
AUTHORS           Ray,L.A., Loh,P., Qi,S., Sloan,D., McDermid,H. and Roe,B.A.
TITLE             Direct Submission
JOURNAL           Submitted (27-MAY-2000) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
COMMENT           On Feb 29, 2000 this sequence version replaced gi:6249691.
                  Because these overlapping clones came from different libraries
                  there are numerous instances of insertions, deletions, and single
                  nucleotide polymorphisms in the overlapping regions below.
                  AC005301(p15j16) 77414 113688 (0) overlaps AC007064(p8708) 1 36275
                  (88548) AC007064(p8708) 52616 124823 (0) overlaps AC006548(p20K14)
                  1 62199 (137950).
FEATURES          Location/Qualifiers
Source            1. 124823
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="22q11"
                  /clone="p8708"
BASE COUNT       41074 a 23932 c 23324 g 36493 t
ORIGIN
Query Match      20.5%; Score 276.4; DB 9; Length 124823;
Best Local Similarity 64.9%; Pred. No. 2.4e-68;
Matches 409; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
QY 601 AGAGTTGTGCTAATGTTATTTCCCTGGTATCTGTCTACCTATGGGCCACCCCTTGCAT 660
Db 24116 ATAGCATGCTGATGACATTTTCCCTGTTATCAGTTACTTATGGGCCCATTCGCTGCAAT 24057
```

Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C.,
Kratochvic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Mosser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokkenkw, S., Oguh, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pul, L., Quiles, M., Renvy, G.,
Rivers, M., Rojas, A., Rojokhan, I., Roife, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Sodergren, E., Sotnik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabof, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vers, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczkyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (Bases 1 to 168918)
Worley, K.C.
Direct Submission
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (Bases 1 to 168918)
Worley, K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:19718450.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTLM
Center clone name: CH230-139K19
----- Summary Statistics
Sequencing vector: Plasmid,
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 104166 bases at least Q40
Consensus quality: 113489 bases at least Q30
Consensus quality: 119214 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1568: contig of 1568 bp in length
* 1569 1668: gap of unknown length
* 1669 2687: contig of 1019 bp in length
* 2688 2787: gap of unknown length
* 2788 4137: contig of 1350 bp in length
* 4138 4237: gap of unknown length
* 4238 5268: contig of 1031 bp in length
* 5269 5368: gap of unknown length
* 5369 6665: contig of 1297 bp in length
* 6666 6765: gap of unknown length

Submitted (09-MAY-2001) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
3 (bases 1 to 160200)
Montgomery, K.T., Grills, G., Li, L., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Haider, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
Direct Submission
Submitted (06-JUL-2002) Harvard Partners Center for Genetics and
Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
02139, USA
4 (bases 1 to 160200)
Montgomery, K.T., Grills, G., Li, L., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Haider, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
Direct Submission
Submitted (02-AUG-2002) Harvard Partners Center for Genetics and
Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
02139, USA
On Aug 2, 2002 this sequence version replaced gi:21700544.
-----Genome Center:
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: <http://www.hpcgc.org/sequence/mouse.html>
Contact: hpgc@mendel.mgh.harvard.edu

[illegible]

Query Match 16.1%; Score 218; DB 2; Length 200698;
Best Local Similarity 57.5%; Pred. No. 2.3e-51;


```
Matches 392; Conservative 0; Mismatches 290; Indels 0; Gaps 0;
y 566 ATGTGAGCTGATCTCTGAGAGGTTCCCTCGGTAGAGTGTGCTAATGGTATTTTCCC 625
b 119410 AGGTGAGATTTCAGAGTCTCTTTCTCTCTCTCTTCAGGCTTCATGACCTTGTCTCC 119469
y 626 TGGTATCTGTACCTATGCGGCGCCACCTTTGCAATATGTTGGCTATCCAGATCAAGTAGC 685
b 119470 TGGTGTCTATTTCTATGCGGCTTACGTTGCAACATCTTAGCCATCAAAATCAAGTAGT 119529
y 686 ATGACTACAGATTCCGCTTGGGCGCACTAGAGTCTCTGCAATCACCATCTCGGCGGACAT 745
b 119530 ATGACTATGAGTCAAGTCAAAACCCCTGGCTTATGTTCTGTATCTTCTCTGGAGAAGCT 119589
y 746 TGGAGATCACTTCCCGCTCTGATTTCTGGTCTCTCTTCAGCCACTTTTGAATTAAGG 805
b 119590 TTGAGATTGCCACCGAGTCATCGTCTGGTCTCTTTACCTCTGCTCTGAAGATCTGGG 119649
y 806 CTGTGCTCTCTAGTGTCTAACTTCCTGATCATCTCTTTGAGCCCTGGATTAAGTTCT 865
b 119650 TGGTGGCAGTCATACCTCGTCACTTCTTCACTTCTTATATCCCTGGATCGTTTCT 119709
y 866 GGAGAAGTGGTGGCCAGATGCCAATACATTGAGAAACCTTCAGCCGCGGTGCGGCACTC 925
b 119710 GTGCGAGTGGCTCCCAATTCCTTGAGAACATAGAGAGGCCCTTAAGTAGGTGGGTACCA 119769
y 926 TGGTGTCTGCTGATTTCACTGTCACCATCTCTATGCTGGCATCAACTTCTCTCTGCTGTCAG 985
b 119770 CCATCGTCTCTGCTCTCTCACTTACTTACTATGCGGATCAACATGTTCTCTGCTGTCAG 119829
y 986 CTTTGGCAGTTGAGTGGCAGACAGAGATCTCGTCGACAAAGGGGAGAACTCGGGGACATA 1045
b 119830 CTGTTTCACTGCTGAAATCGAATCCGGAGCTCATCAGCAAGTCCCGAAGCTGGTACCGCC 119889
y 1046 TGGCCCTGCATATAGTGTGAGTGGTAGAGATGATCATGCTCTGTTGTTTAAAGT 1105
b 119890 TGCTCATTTACTATGATGACGAGATTCATCGAAGTCCGCTCTCTCTCTCTGCTGCTGAT 119949
y 1106 TCTTTGAGTGAAGTGTACTGAATTAATGCTTCTTCTGATTCCTTGGAGCTTCATTA 1165
b 119950 TTTTAAACTGACATATACATGATGATGTTGCGCCCTCTGTTGATCTTGACGCTGCTTA 120009
y 1166 TTGCTTATCTGATTTCCATTCATCTCATGCTCTCTTTCTTCCAGTACTTGCATCCATTCG 1225
b 120010 TTGGTACTGCACTGGCACTTCTTCTCATGCTGTGTTCTATCATGTTTTTCCACCTTCCA 120069
y 1226 GTCACATCTTCAACCATTAATCT 1247
b 120070 AAAAGCTCTTCTCTCCAGTGT 120091
```

```
RESULT 14
XUS :016752/c
DEFINITION Homo sapiens BAC clone RP11-506M9 from Y, complete sequence.
AC016752
XUS :016752.2 GI:7637819
KEYWORDS HTG.
XUS :
DEFINITION Homo sapiens
XUS :
ORGANISM Homo sapiens
XUS :
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166436)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
1 (bases 1 to 166436)
Nguyen, C., Maupin, R., Hawkins, M. and Smith, R.
The sequence of Homo sapiens BAC clone RP11-506M9
Unpublished
3 (bases 1 to 166436)
```

```
AUTHORS
TITLE
JOURNAL
Waterston, R.H.
Direct Submission
Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 166436)
Waterston, R.H.
Direct Submission
Submitted (24-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 166436)
Waterston, R.H.
Direct Submission
Submitted (24-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 166436)
Waterston, R.H.
Direct Submission
Submitted (25-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
7 (bases 1 to 166436)
Waterston, R.H.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 23, 2000 this sequence version replaced gi:6524293.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
-----
Summary Statistics
Center project name: H_NH0506M09
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-221K4. Actual start of this clone is at base position 1 of RP11-506M9; actual end is at base position 166436 of RP11-506M9.

Accession	Sequence	Length	Source	Organism
113928	GATATATATATTGATATTATTTGCTTGGAGGAAATTTTCACCTGAAGCTTTGTGACTC	113869		
1143	CTTGATGCCCTTGACGACTCATTTATGCTTATCTGATTTCCATTGCACTGCCTCTTTT	1202		
113868	ATTAATGCCATGTAGTTTTCATATAACCACTATTAGCCATTGCTTTATGCTCCTCTT	113809		
1203	CTTCAGTACTTGCACTCCATTGCGCTCA	1230		
113808	CTGCAGTATTGTGCACCAAGTGGTCA	113781		
SULT 15				
017005	AC017005	161879 bp	DNA	linear
FINITION	Homo sapiens BAC clone RP11-100J21	from Y,	complete	sequence.
CESSION	AC017005			
RSION	AC017005.6	GI:13435269		
YWORDS	HTG.			
URCE	Homo sapiens.			
ORGANISM	Homo sapiens			
PERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
JOURNAL	1. (bases 1 to 161879)			
TITLE	Sulston, J.E. and Waterston, R.			
MEDLINE	Toward a complete human genome sequence			
PUBMED	Genome Res. 8 (11), 1097-1108 (1998)			
PERENCE	99063792			
AUTHORS	9847074			
JOURNAL	2. (bases 1 to 161879)			
TITLE	Armstrong, J., Stoneking, T., Hawkins, M. and Bernard, K.			
PERENCE	The sequence of Homo sapiens BAC clone RP11-100J21			
AUTHORS	Unpublished			
JOURNAL	3. (bases 1 to 161879)			
TITLE	Waterston, R.H.			
PERENCE	Direct Submission			
AUTHORS	Submitted (09-DEC-1999)			
JOURNAL	Genome Sequencing Center, Washington			
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
PERENCE	MO 63108, USA			
AUTHORS	4. (bases 1 to 161879)			
JOURNAL	Waterston, R.H.			
TITLE	Direct Submission			
PERENCE	Submitted (23-MAR-2001)			
AUTHORS	Genome Sequencing Center, Washington			
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
TITLE	MO 63108, USA			
PERENCE	5. (bases 1 to 161879)			
AUTHORS	Waterston, R.			
JOURNAL	Direct Submission			
TITLE	Submitted (09-MAY-2001)			
PERENCE	Department of Genetics, Washington			
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA			
JOURNAL	On Mar 23, 2001 this sequence version replaced gi:19211335.			
COMMENT	----- Genome Center			
	Center: Washington University Genome Sequencing Center			
	Center code: WUGSC			
	Web site: http://genome.wustl.edu/gsc			
	Contact: sapiens@watson.wustl.edu			
	----- Summary Statistics			
	Center project name: H NH0100J21			

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaltsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:

The RPl-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Pengen, E., Tateo, M., Catanesi, J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is Rp11-79J10; the clone sequenced to the right is Rp11-24SK4, 200 bp overlap. Actual start of this clone is at base position 1 of Rp11-100J21; actual end is at base position 66964 of Rp11-24SK4.

FEATURES

```

1. .161879
/organisms="Homo sapiens"
/db xref="taxon:9606"
/chromosome="Y"
/map="Y"
/clone="RP11-100321"
/clone_lib="RPC1-11"
1. .1185
/rpt_family="L1"
1183. .1513
/rpt_family="L1"
1679. .1842
/rpt_family="WERV1_type"
2125. .3135
/rpt_family="ERV1"
3116. .3412
/rpt_family="ERV1"
3450. .3866
/rpt_family="ERV1"
3867. .6977
/rpt_family="ERV1"
6986. .7295
/rpt_family="ERV1"
7297. .7625
/rpt_family="ERV1"
9507. .9570
/rpt_family="WERV1_type"
11143. .11817
/rpt_family="L1"
12324. .12843
/rpt_family="L1"
12874. .12922
/rpt_family="ERV1"
12885. .13929
/rpt_family="ERV1"
13051. .13264
/rpt_family="L1"
13361. .13619
/rpt_family="L1"
13628. .14306
/rpt_family="L1"
15532. .15662
/rpt_family="ERV1"
15740. .16044
/rpt_family="ERV1"
16045. .16136
/rpt_family="ERV1"
16178. .16239
/rpt_family="ERV1"
16728. .17201
/rpt_family="ERV1"

```

BEST AVAILABLE COPY

```
repeat_region /rpt_family="ERV1"
repeat_region 17188..17253
repeat_region /rpt_family="ERV1"
repeat_region 17243..17602
repeat_region /rpt_family="ERV1"
repeat_region 17746..18064
repeat_region /rpt_family="L1"
repeat_region 18085..18235
repeat_region /rpt_family="L1"
repeat_region 18303..18876
repeat_region /rpt_family="ERVK"
repeat_region 18927..19413
repeat_region /rpt_family="ERVK"
repeat_region 19419..19628
repeat_region /rpt_family="L1"
repeat_region 21508..21759
repeat_region /rpt_family="Alu"
repeat_region 22725..22868
repeat_region /rpt_family="ERV1"
repeat_region 22868..23335
repeat_region /rpt_family="ERV1"
repeat_region 23350..24076
repeat_region /rpt_family="ERV1"
repeat_region 24273..24690
repeat_region /rpt_family="ERV1"
repeat_region 24702..24823
repeat_region /rpt_family="ERV1"
repeat_region 24886..25175
repeat_region /rpt_family="Alu"
repeat_region 25186..27543
repeat_region /rpt_family="ERV1"
repeat_region 27654..27697
repeat_region /rpt_family="ERV1"
repeat_region 28186..28421
repeat_region /rpt_family="ERV1"
repeat_region 28440..28804
repeat_region /rpt_family="ERV1"
repeat_region 28773..29267
repeat_region /rpt_family="ERV1"
repeat_region 29270..31939
repeat_region /rpt_family="ERV1"
repeat_region 31933..32211
repeat_region /rpt_family="ERV1"
repeat_region 32225..32546
repeat_region /rpt_family="Alu"
repeat_region 32582..33624
repeat_region /rpt_family="ERV1"
repeat_region 33740..34049
repeat_region /rpt_family="ERV1"
repeat_region 34050..34149
repeat_region /rpt_family="ERV1"
repeat_region 34155..34617
repeat_region /rpt_family="ERV1"
repeat_region 34624..37556
repeat_region /rpt_family="ERV1"
repeat_region 37676..37979
repeat_region /rpt_family="ERV1"
repeat_region 38072..38379
repeat_region /rpt_family="ERV1"
repeat_region 38380..39312
repeat_region /rpt_family="ERV1"
repeat_region 39318..40245
repeat_region /rpt_family="ERV1"
repeat_region 40372..40443
repeat_region /rpt_family="ERV1"
repeat_region 40486..41009
repeat_region /rpt_family="ERV1"
```

Query Match 15.9%; Score 214; DB 9; Length 161879;
Best Local Similarity 61.9%; Pred. No. 3.3e-50;
Matches 389; Conservative 0; Mismatches 235; Indels 4; Gaps 3;
QY 603 AGTTGCTAATGTTATTTCCCTGGTATCTGTCCACCTATGGGGCCACCCTTTGCAATAT 662

```
Db 160312 AGCATTTGCTGATGACATGTTCTCTTTACCACTTGTGTATGGGCCATTCTACGCAATAC 160371
QY 663 GTTGGCTATCCAGATCAAGTAGCATGACTACAAGATTTCGGCTTTGGGCCACTAGAAATCCT 722
Db 160372 ACTGGCCATCCATACCAGCAACGACTTTCTCTGACATTTAGCTACAAGCGATAAATTCAT 160431
QY 723 CTGATCAACATCTTGGCGGACATTTGAGATCACTTCCGGCTCTCTGATTTCTGTTGCTCTT 782
Db 160432 CTG--TCATGATACGGTGTCTCTTGGTATTTATCTCACGTGTAGTACTCTGGAATTTT 160489
QY 783 CTCAGCCACTTTGAATTTGAAGCTGTGCCCTTCCCTTAGTCTCAACTTCTCTGATCATCT 842
Db 160490 TCTGTCATCTCTGAAACAGAGGAGCCTACCAITTAATTAATCATATATTTTGTATTATT 160549
QY 843 CTTTGAAGCCCTGGATTAAGTTCTTGGAGAGTGGTGGCCAGATGCCCAATTAACATTGAGAA 902
Db 160550 GTTGGCACCATGGCTGGAGTTTTTGGAAAGTGGAGCTCATCTTTCATAACAACAGAGAAA 160609
QY 903 AAACCTTCAAGCGGGTCCGGCACTCTGGTGGTCTCTGATTTTCACTCACCATCTCTATGCTGG 962
Db 160610 TAATTCAGCATGGTGGGTAC-ATACGGATGCTTACCTTAATCATGCTACTATATGCTGC 160668
QY 963 CATCAACTTCTCTTGGTGGTCACTTTGAGTTGAGTTGGCAGACAGAGATCTCGTCA 1022
Db 160669 TATCAAACTTCTCTGGCTGGTCACTGAAACTGCACTCTCAATGAGGAAGTAATGA 160728
QY 1023 CAAAGGGCAGAACTGGGGACATATGGCCCTGCACATAGTGTGAGGTTGGTAGAAGATGT 1082
Db 160729 CAAGAGACTGAGGTGGGC-CATAGAATCCTACACTACAGCTTCTAGTTTTTAGAAGATGT 160787
QY 1083 GATCATGGTCTTGGTTTTTAAAGTTCTTTGGAGTGAAGTGTACTGAATTAAGTCTATTC 1142
Db 160788 GATAATAATATTGATATTTATGTTTGGTGGAGGAAAAATTCACGAAAGTCTTGTGACTC 160847
QY 1143 CTTGATTGCTTGCAGCTCATTTATGCTTATCTGATTTCCATTGACCTCATGCTCCTTTT 1202
Db 160848 ATTAATGGCATGTAGTTCATCAATAACCTTAACCTATTAGCCATTGGCTTTATGCTCTCTT 160907
QY 1203 CTTCCAGTACTTGCATCCATTGGGCTCA 1230
Db 160908 CTGTCAGTATTTGCACCCCAAGGTGGTCA 160935
```

Search completed: March 30, 2003, 05:11:04
Job time : 3911.69 secs